High-fidelity, efficient, and reversible labeling of endogenous proteins using CRISPR-based designer exon insertion

Haining Zhong^{1, *}, Crystian I. Massengill¹, Michael A. Muniak¹, Lei Ma¹, Maozhen Qin¹, Stefanie Kaech Petrie², and Tianyi Mao¹

1. Vollum Institute, Oregon Health & Science University, Portland, Oregon, 97239

2. Department of Neurology, Oregon Health & Science University, Portland, Oregon, 97239

*: Correspondence should be addressed to H.Z. (email: zhong@ohsu.edu).

Haining Zhong

Vollum Institute, Oregon Health & Science University

3181 SW Sam Jackson Park Rd., L474

Portland, Oregon 97239

U.S.A.

zhong@ohsu.edu / 503-494-5089

ABSTRACT

Precise and efficient insertion of large DNA fragments into somatic cells using gene editing technologies to label or modify endogenous proteins remains challenging. Non-specific insertions/deletions (INDELs) resulting from the non-homologous end joining pathway make the process error-prone. Further, the insert is not readily removable. Here, we describe a method called CRISPR-mediated insertion of exon (CRISPIE) that can precisely and reversibly label endogenous proteins using CRISPR/Cas9-based editing. CRISPIE inserts a designer donor module, which consists of an exon encoding the protein sequence flanked by intron sequences, into an intronic location in the target gene. INDELs at the insertion junction will be spliced out, leaving mRNAs nearly error-free. We used CRISPIE to fluorescently label endogenous proteins in neurons *in vivo* with previously unachieved efficiency. We demonstrate that this method is broadly applicable, and that the insert can be readily removed later. CRISPIE permits protein sequence insertion with high fidelity, efficiency, and flexibility.

INTRODUCTION

The CRISPR/Cas9 technology has revolutionized genomic editing (Cong et al., 2013; Doudna, 2020; Heidenreich and Zhang, 2016; Jinek et al., 2012; Mali et al., 2013). However, one of the major unresolved challenges is to use CRISPR-based technologies to precisely and efficiently knock in large DNA fragments in somatic cells of living animals (i.e., without generating transgenic animals) to label or modify endogenous proteins. Such technology holds promise for both studying biological mechanisms and gene therapy.

Towards this goal, several studies have developed CRISPR/Cas9-based approaches to insert fluorescent protein (FP)-encoding sequences to label and visualize endogenous proteins in somatic cells, such as postmitotic neurons (Gao et al., 2019; Mikuni et al., 2016; Nakade et al., 2014; Nishiyama et al., 2017; Schmid-Burgk et al., 2016; Suzuki et al., 2016; Uemura et al., 2016; Willems et al., 2020). Fluorescent visualization of proteins in living cells is essential for the mechanistic dissection of cellular and organismic processes because many functions of cells and organisms are established and manifested by their constituent proteins. However, conventional methods have fallen short in this regard: immunolabeling is typically incompatible with live imaging, and it usually does not permit sparse labeling for cell-specific contrast in tissues. FP tagging typically involves overexpression, which may result in undesired off-target effects. Knock-ins of FP tags are costly and time-consuming in mammalian species, and, with the exception of certain complex schemes (e.g., Fortin et al., 2014; He and Huang, 2018), are typically associated with poor contrast due to global expression of the FP-tagged protein of interest (e.g., Herzog et al., 2011). In contrast, if successful, CRISPR-based FP labeling of

endogenous proteins in somatic cells may overcome these limitations and remove a major bottleneck in studying biological mechanisms.

Unfortunately, current CRISPR-based somatic protein labeling technologies are faced with major limitations. First, the *in vivo* labeling efficiency is moderate especially for the large DNA insertions required for FP labeling (~15% or less) (Mikuni et al., 2016; Nishiyama et al., 2017; Suzuki et al., 2016; Uemura et al., 2016). Second, and more importantly, these methods, which target protein coding sequences or their immediately adjacent exonic sequences, are error-prone. The precise insertion of short (< 50 bp) DNA sequences has started to become possible via prime editing (Anzalone et al., 2019), which only nicks the DNA; however, the insertion of larger DNA fragments still requires the generation of double-strand DNA breaks (DSBs). At the edited loci involving DSBs, including those loci where the DNA insertion does not occur, the efficient NHEJ pathway often results in unwanted INDELs that cause mutations and/or frame shifts (Cong et al., 2013; Doudna, 2020; Heidenreich and Zhang, 2016; Mali et al., 2013) (See also Figure 2 and Figure 1—figure supplement 1). This is the case regardless of whether NHEJ or homologous-dependent repair (HDR) is engaged for the insertion process (Roberts et al., 2017). Finally, it has been increasingly recognized that reversible DNA editing may allow for great flexibility in both research and gene therapy (Nakamura et al., 2019); however, the inserted labels cannot be readily removed at a later time using existing approaches.

Here, we present a strategy, called CRISPR-mediated insertion of exon (CRISPIE), which allows for the nearly error-free insertion of FP sequences with high efficiency. Instead of targeting

gene exons, CRISPIE targets introns and inserts a designer donor module, which includes an exon encoding the desired protein sequence and the surrounding intronic sequences. INDELs occurring at the insertion junction within the intronic region of DNA will be spliced out (Figure 1A), resulting in very low error rates at the mRNA level (>98% correct). CRISPIE is flexible and broadly compatible with: 1) both N- and C-terminal labeling, 2) proteins with diverse structures and functions, including pre- and post-synaptic proteins and cytoskeletal proteins, 3) all major transfection methods, 4) FPs with diverse colors, and 5) multiple animal species. In part because introns offer ample editing sites to choose from, and because INDELs at the DNA level do not affect the success of editing, a high labeling efficiency (up to 43%) was achieved in cortical neurons of living mice. Importantly, CRISPIE-mediated DNA insertions are erasable. By flanking the donor module with additional designer CRISPR editing sites in the intronic region, the inserted DNA fragment can be erased at a later time. CRISPIE may allow for the routine labeling of proteins at endogenous levels, and can be expanded to the insertion of other genetically encoded functional sequences to manipulate protein function.

RESULTS

Design and demonstration of CRISPIE

To achieve high-fidelity FP labeling, we developed the CRISPIE method, which inserts a designer exon module that consists of an exogenous exon encoding the FP flanked by intronic sequences, including the splicing acceptor and donor sites, into the intronic region of the target gene via the NHEJ pathway (Figures 1A, *cf.* Figure 1–figure supplement 1). NHEJ exhibits a higher editing efficiency than homologous dependent repair (HDR) in postmitotic somatic cells,

such as neurons (Heidenreich and Zhang, 2016; Heyer et al., 2010; Mao et al., 2008; Saleh-Gohari and Helleday, 2004). Because introns are spliced out of mRNAs, they are often poorly conserved (e.g., see Figure 1–figure supplement 2) and tolerate stochastic mutations better than exons. INDELs at the joint junction of the insertion therefore will not result in mutations in the mRNA and thus the encoded protein under most circumstances (Figure 1A). In addition, NHEJ-mediated donor DNA insertion can result in unwanted integration with inverted orientation, and INDELs can occur at loci where the donor fails to insert. However, neither of these events will produce disrupted mRNAs when using CRISPIE (Figure 1A). Conceptually, unlike conventional exonic targeting, CRISPIE only results in either wild-type or successfully labeled mRNAs. For convenience, after a gene is labeled by CRISPIE, we say it has been “CRISPIEd”.

As a proof of principle, we first labeled human β -actin (*hACTB*) at its N terminus with monomeric EGFP (mEGFP) in human U2OS cells by co-transfecting a plasmid that expresses a single guide RNA (sgRNA) targeting the *hACTB* gene and *Streptococcus pyogenes* Cas9 (SpCas9), a plasmid carrying the donor module (Figure 1B and Figure 1–figure supplement 3B1), and a transfection marker plasmid that expresses mRuby3. The sgRNA/SpCas9 complex cuts an editing site at the first intron of *hACTB* (Figure 1C), as selected using the BROAD Institute sgRNA designer (Doench et al., 2016), and releases the donor module from the plasmid by cutting sgRNA-targeted sequences flanking the module (Figure 1B, lower panel). The released module is then inserted into the cut site at the *hACTB* intron via the NHEJ pathway. The module-flanking sgRNA-targeted sequences are in the reverse orientation to facilitate the insertion of the module in the forward orientation (Suzuki et al., 2016). The donor module includes an exon in

the appropriate translational phase (0-0 for intron 1 of *hACTB*) encoding mEGFP and peptide linkers (Figure 1–figure supplement 3B1). At the exon-intron junctions, ~100 bp of intronic sequence and ~10-bp of adjacent exonic sequence were taken from an obligatory intron (i.e., no reported alternative splicing events) and its adjacent exons of mouse CaMKII α (*mCamk2a*), a highly expressed protein in the brain, to include the splicing donor or acceptor site (Wang and Burge, 2008). mEGFP labeling was observed at 3 – 4 days post-transfection (Figure 1D), and required the presence of both sgRNA/SpCas9 and the donor module. Over 35% ($37 \pm 5\%$) of transfected cells, as identified by the expression of mRuby3 (red), were mEGFP (green)-labeled (Figure 1E). This likely represents a lower bound of labeling efficiency, since the co-transfection rate of all three plasmids is unlikely to reach 100%. Confocal and super-resolution structured illumination microscopy (SIM) revealed that the mEGFP signals exhibited the characteristic distribution of actin protein (Fischer et al., 1998; Planchon et al., 2011) (Figure 1F). Time-lapse confocal microscopy revealed dynamic actin ruffles at the edges of cells (Figure 1G), which is consistent with the current knowledge of actin dynamics (Fischer et al., 1998; Planchon et al., 2011) and suggests that mEGFP-labeled actin retains the functions of the wild-type protein.

CRISPIE-mediated protein insertion is nearly error-free at the mRNA level

To characterize the insertion events and INDEL rates at the insertion site, we sorted transfected cells into red-only (transfected, but not labeled) and green (transfected and labeled) cells. The actin genomic DNA and mRNA at the editing locus were analyzed using PCR and RT-PCR, respectively (Figures 2A and 2B; see Figure 2–figure supplement 1 and Supplementary Tables 1 and 2 for primers, and anticipated PCR products and sizes). As expected, bands corresponding to the successful insertion (i.e., insertion in the forward

orientation) at the targeted site are observed in green-labeled cells for both the genomic DNA and the mRNA (Figure 2A). Systematic analyses of green-labeled cells detected all possible editing events at the genomic DNA level, including no label insertion, forward insertion, and inverted insertion (Figure 2B), indicating that insertion events were typically non-homologous (i.e., both chromosomes differentially edited). Had the coding sequence been targeted, an inverted label insertion would inevitably cause disruptive mutations. However, under our conditions, only wild-type and forwardly inserted mRNAs were detected (Figure 2B), demonstrating the advantage of CRISPIE.

To evaluate how INDELs may affect a targeted gene, the PCR amplicons corresponding to no label insertion events in untransfected, red-only, and green cells, as well as those corresponding to forward insertion events in green cells were subjected to next-generation sequencing (NGS) (Figures 2C and 2D; sequenced amplicons correspond to gel bands in Figures 2A and 2B, as indicated by Roman numerals). Remarkably, at the genomic DNA level in red-only cells, over 70% (834/1181 reads) of PCR amplicon sequences contained INDELs (Figure 2C, and 2nd column of Figure 2D), confirming the predicted high rate of INDELs during gene editing. In contrast, at the mRNA level in these cells, 99.7% (2930/2936 reads) of the sequences did not contain INDELs ($p < 0.001$, cf. genomic DNA) (Figure 2D, 2nd column), which was no different from control, untransfected cells (99.7%, 5761/5771 reads) (Figure 2D, 1st column). Similarly, undesired INDELs were frequently detected in all sequenced amplicons from green cells at the genomic DNA level, but few were detected at the mRNA level (INDELs 24 – 87% for DNA vs 0.3 – 2% for mRNA; Figure 2D). These numbers correspond to correct-to-mutation ratios of 50 – 300 fold for mRNA, which are around two orders of magnitude higher than those for DNA. These

results demonstrate that, although inverted insertion and INDEL events may occur at the DNA level — which would have caused mutations or frame shifts when using existing strategies to target coding sequences — CRISPIE is resistant to these disruptive events at the mRNA level.

CRISPIE can be optimized to achieve high labeling efficiency

For insertion strategies that target coding sequences, the options for targeting sites are limited, and not all potential targeting sites allow for high-efficiency editing (Doench et al., 2014). Furthermore, as discussed above, INDELs cause mutations or frame shifts, thereby reducing the likelihood of successful labeling. In contrast, CRISPIE is insensitive to INDELs, and the lengths of introns usually offer ample choices for potential targeting sites. CRISPIE may therefore have the potential to achieve higher labeling efficiency than strategies that target coding sequences. To test this, we first asked whether different intronic insertion sites may exhibit different labeling efficiencies. Five different editing sites (i1 to i5 in Figure 3A) were selected using the BROAD Institute sgRNA designer (with Azimuth v2.0 scores of 0.57–0.69) (Doench et al., 2016). By using a generic donor that was released from the plasmid by a designed sgRNA (called DRS-2; see Figure 1–figure supplement 3B7 for donor plasmid design) (Gao et al., 2019), we found that all five sites permitted the successful labeling of *hACTB*; however, the labeling efficiencies varied significantly (Figure 3B), with the highest labeling efficiency corresponding to around 20-fold of the lowest.

We also examined FP labeling that directly targets the coding sequence for comparison. There are only two possible SpCas9 editing sites around the start codon of *hACTB* (referred to as e1 and e2; Figure 3A). Despite using their optimal insertion donors (for both, see Figure 1–

figure supplement 3B6), both exonic sites exhibited significantly lower labeling efficiencies compared to the optimized CRISPIE editing ($29 \pm 4\%$ and $5 \pm 1\%$ of the best CRISPIE efficiency for e1 and e2, respectively; for both $p < 0.001$) (Figure 3C). Between the e1 and e2 sites, the one with the higher Azimuth score was in fact less efficient (Azimuth scores = 0.55 and 0.62, respectively), indicating that the labeling efficiency is not yet entirely predictable. Nevertheless, CRISPIE allows for systematic optimization by screening for highly efficient editing sites and, when optimized, can achieve higher labeling efficiency than exonic editing.

CRISPIE is broadly applicable in dividing cells

The CRISPIE method should be broadly applicable for use with different FPs or with other functional domains, different protein targets, and different animal species. To test different FP labels for potential multiplex imaging studies, we successfully labeled hACTB in U2OS cells using cyan (mTurquoise2), green-yellow (mNeonGreen), and red (mRuby3) FPs (Figure 4A, donor plasmids illustrated in Figure 1–figure supplement 3B2 – 3B4). To test the applicability of CRISPIE to different protein targets, we labeled several additional important subcellular structures in U2OS cells, namely microtubules, mitochondria, focal adhesion complexes, and the endoplasmic reticulum (ER), by labeling tubulin alpha 1B (hTUBA1B), translocase of outer mitochondrial membrane 20 (hTOMM20), vinculin (hVCL), and calreticulin (hCALR), respectively (Figure 4B and Figure 1–figure supplement 3A and 3B) (Lam et al., 2012; Rizzo et al., 2009; Roberts et al., 2017; Shroff et al., 2008). Notably, the *hTOMM20* and *hVCL* genes were labeled at their respective last introns (Figure 1–figure supplement 3A) because their encoded proteins are best tagged at their C termini (Rizzo et al., 2009; Roberts et al., 2017; Shroff et al., 2008). This demonstrates that CRISPIE is also readily compatible with C-

terminal labeling. For some genes, such as *hVCL*, there are coding sequences of significant lengths present at the last exon, downstream of the possible CRISPIE labeling site. To prevent potential disruption of protein function caused by an insertion in the middle of the protein, the coding sequence of the last exon of *hVCL* was included in the insertion donor sequence, and a stop codon was introduced at the end of the mEGFP coding sequence (Figure 1–figure supplement 3B5). To test the applicability in different animal species, we further labeled ACTB (mACTB) in mouse Neuro 2A cells. A mouse-specific sgRNA was used because introns are degenerative across species (e.g., see Figure 1–figure supplement 2). The labeling was successful and was dependent on the correct combination of the species and the corresponding sgRNA (Figure 4C). CRISPIE labeling was also successful in rat neurons (Figure 5A, see below). Overall, these results indicate that CRISPIE is versatile and readily applicable to many different FP labels, protein species, and animal species, and can be used to tag proteins at their N and C termini.

CRISPIE is broadly applicable to postmitotic neurons using various transfection methods *in vitro* and *in vivo*

To determine whether CRISPIE can be used in postmitotic cells, such as neurons, and whether it is compatible with different transfection methods, we applied it to label proteins in neurons in cultured slices and *in vivo* using three common transfection methods. First, we used the biolistic method (i.e., “gene gun”) to deliver the sgRNA and donor for rat β -actin (rACTB), and the cytosolic marker mRuby3, to cultured rat hippocampal slices. Notably, although the biolistic transfection method is widely used in studying the cellular functions of neurons, CRISPR-based labeling has not yet been demonstrated using this method, likely due in part to

the relatively low labeling efficiencies. For this same reason, high-contrast double-labeling — e.g., the target protein together with a cell morphology marker — in brain tissue remains challenging. In our experiments, 6 – 45 neurons were transfected per brain slice, as determined by their red mRuby3 fluorescence. Among them, on average 15% ($15 \pm 4\%$, $n = 6$ slices) exhibited successful labeling of actin (Figure 5A). In some slices, as many as 6 – 7 successfully labeled neurons of different neuronal types, including dentate gyrus neurons, CA1 and CA3 pyramidal neurons, and interneurons, could be detected (Figure 5—figure supplement 1). Glial cells could also be labeled (Figure 5—figure supplement 1B). These results indicate that CRISPIE provides a viable option for protein labeling using gene gun in cultured slices. Similar results were obtained in cultured mouse hippocampal slices from a *SpCas9* heterozygous genetic background, which constitutively expresses SpCas9 and thereby simplified the experimental design (Figure 5B). Other synaptic proteins, including synapsin (SYN1), tubulin (TUBA1B), and CaMKII α , were also successfully labeled (Figures 5A and 5B). In many labeled cells, both the mRuby3 cytosolic marker and mEGFP were readily detectable, allowing us to compare the subcellular distribution of three proteins — β -actin, α -tubulin 1B, and CaMKII α — at neuronal dendrites (Figure 5C). By quantifying their distribution between the dendrites and spines using a previously described spine enrichment index (SEI) (Zhong et al., 2009), we found that the three proteins exhibited distinct subcellular distributions: actin was highly enriched in dendritic spines, and CaMKII α was moderately enriched, whereas tubulin was excluded from spines (Figure 5C; see Star Methods for quantification details).

To determine whether CRISPIE can be used to label endogenous proteins *in vivo*, *in utero* electroporation (IUE) was used to introduce the sgRNA, the donor, and the cytosolic

marker mRuby3 into the somatosensory cortex of *SpCas9* mice to label β -actin or tubulin in separate experiments (Figure 5D). Among the transfected red cells, the labeling efficiencies were $43 \pm 5\%$ for actin and $30 \pm 5\%$ for tubulin (Figure 5E), both of which are significantly higher than previously reported for CRISPR-based endogenous labeling of any protein *in vivo* (Mikuni et al., 2016; Nishiyama et al., 2017; Suzuki et al., 2016; Uemura et al., 2016). This high level of labeling efficiency allows us to routinely perform *in vivo* imaging in the brains of living mice (Figure 5F and Figure 5—figure supplement 2). Finally, as viruses have become an increasingly prevalent method of transfection, we generated an adeno-associated virus (AAV) harboring the sgRNA targeting β -actin, the donor, and mRuby3 (Figure 1—figure supplement 3C). When injected into the cortex of *SpCas9* mice, this virus successfully transduced neurons *in vivo* and labeled the actin cytoskeleton (Figure 5G). Overall, CRISPIE can be readily applied using a variety of widely used methods to transfect neurons both *in vitro* and *in vivo* with high efficiency.

CRISPIE can be reversible

Reversible gene editing may offer significant benefits in research, and in gene therapy in the future. However, existing CRISPR-based methods for endogenous labeling and large-fragment DNA insertions are not readily reversible. CRISPIE allows for the inclusion of functional elements at the flanking intronic regions of the donor module without affecting the mRNA. Taking advantage of this, we included additional designer sgRNA targeting sequences (DRS-1; Gao et al., 2019) at both ends of the module (Figure 6A). These added sites allowed for the later excision of the inserted module guided by the DRS-1 sgRNA. To demonstrate this reversibility, we isolated a U2OS cell clone of CRISPIEd hACTB. All cells derived from this clone uniformly

exhibited green-labeled actin structures (Figure 6B, left panel). When DRS-1 sgRNA and SpCas9 were co-transfected with the transfection marker mRuby3, the green label was absent or greatly attenuated in over three quarters ($76 \pm 6\%$) of the mRuby3-positive cells at 5 days after transfection (Figures 6B and 6C). As a control, transfection with an sgRNA targeting mACTB did not result in detectable loss of the green labeling. To the best of our knowledge, this is the first demonstration of a readily reversible gene editing approach at the DNA level.

DISCUSSION

Here, we present CRISPIE, a reversible CRISPR/Cas9-based approach for inserting large DNA fragments to label or modify endogenous proteins that is essentially impervious to the side-effects of INDELS. CRISPIE achieves a high level of precision by inserting a designer exon module flanked by adjacent introns into an intronic site of the target gene. While intronic editing has been used to introduce non-coding DNA tags, microRNAs, gene disruptions, and exon replacements (Chen et al., 2018; Lee et al., 2018; Li et al., 2016; Miura et al., 2015), all of these studies involved cloning of the cell or organism. Although CRISPIE can conceivably be used when cloning of the edited cell or organism is involved, its strength lies in its applicability to somatic cells. Somatic cell editing offers significant advantages compared to cloning-based approaches, including lower cost, faster turnaround time, and controllable alterations in labeling density. In particular, with regard to labeling density, sparse labeling is desired for high-contrast imaging, whereas high-density labeling is needed for biochemical or proteomic analyses. In addition, somatic cell editing will be required under most circumstances when gene editing is applied directly on patients for therapeutic purposes.

As mentioned earlier, several previous studies have labeled endogenous proteins in somatic cells (Gao et al., 2019; Mikuni et al., 2016; Nishiyama et al., 2017; Schmid-Burgk et al., 2016; Suzuki et al., 2016; Uemura et al., 2016; Willems et al., 2020). However, these methods are sensitive to INDELs because they directly target exons. Notably, when these methods are applied *in vivo*, the majority of transfected cells are not labeled (labeling efficiency $\leq 15\%$), but INDELs can still occur and cause major disruption of the targeted protein in these unlabeled cells. In the (v)SLENDR approach (Mikuni et al., 2016; Nishiyama et al., 2017), which engages HDR as the insertion mechanism, the sequences within a short range before the start codon or after the stop codon can be used as target sites, thereby alleviating the side-effects of INDELs. However, INDELs often include deletions tens of base pairs or longer (e.g., Figure 2C). For many genes, INDELs in the 5' and 3' untranslated regions may potentially affect the trafficking, stability, or translational control of the mRNA (Fortin et al., 2014; Goldie and Cairns, 2012). Furthermore, HDR occurs with low efficiency in non-dividing cells. The vSLENDR approach increases HDR efficiency in postmitotic neurons using high-titer AAV; however, AAV construction and generation add significantly to the cost and time of experiment, and the *in vivo* labeling efficiency remains low ($\leq 15\%$). Finally, the recently developed prime editing technique, while exhibiting high precision, can only insert short sequences (< 50 bp) and has not been demonstrated *in vivo*.

Compared to previous methods, CRISPIE is nearly error-free in both labeled cells, and transfected but unlabeled cells. It is broadly compatible *in vitro* and *in vivo* with all tested transfection methods, including liposome-based transfection, biolistic transfection, in utero electroporation, and viral infection. CRISPIE exhibits approximately 4-fold higher labeling

efficiency in side-by-side comparisons with an exon targeting strategy, and achieves up to 2.7-fold higher efficiency than the best *in vivo* labeling efficiency described in the literature (Nishiyama et al., 2017; Suzuki et al., 2016). This higher labeling efficiency is significant as it greatly increases the feasibility of *in vivo* multiplexed imaging of the labeled protein with the cellular context revealed by a cytosolic marker (e.g., Figure 5F). It also provides a new starting point for further improving the labeling efficiency towards 100%, which may be required for future therapeutic purposes.

CRISPIE is limited by the availability of introns at or near the desired targeting sites. However, as demonstrated, such sites are available in the genes of a large number of protein species for labeling at either the N or C termini. In addition, labeling at an internal site of a gene is feasible as long as the insertion does not disrupt the function of the encoded protein. Many introns reside at the junctions of functional domains because introns have evolved in part to facilitate functional domain exchanges (Kaessmann et al., 2002; Patthy, 1999). These junctions may serve as candidate internal insertion sites for FP labeling. At the same time, although CRISPIE improves upon the precision at the target site, it does not prevent potential off-target cleavages of sgRNA/Cas9. However, increasingly powerful bioinformatics tools allow for the selection of targeting sites with less likely off target effects (e.g., Doench et al., 2014, 2016), and this concern will be further alleviated or eliminated in the future by the continued development of high-fidelity Cas9 variants (Chen et al., 2017; Kleinstiver et al., 2016; Slaymaker et al., 2016). Similarly, although CRISPIE enables the tagging of endogenous proteins with low error rates, it does not ensure that the tagged protein functions the same as the wild-type protein. Not all tagging is benign, and rigorous characterizations will be needed for each tagging experiment.

CRISPIE-based DNA insertion is to our knowledge the first readily reversible gene editing strategy that operates at the DNA level, which offers flexibility in both research studies and in gene therapy in the future. Although the original editing site will be destroyed during the labeling-erasing cycle, adjacent editing sites within the intron can allow for additional rounds of labeling, thereby simulating a reversible process. Overall, CRISPIE can label endogenous proteins with high fidelity, high efficiency, and reversibility in diverse animal species. This method can also conceivably be used to introduce other functional domains to modify protein function for research and for therapeutic purposes. This method may have broad applications in both biomedical research and in medicine.

MATERIALS AND METHODS

Plasmid constructs

Plasmid constructs were generated using gene synthesis or standard subcloning methods. All previously unpublished constructs and their corresponding sequences will be deposited to Addgene upon acceptance of the manuscript.

For most experiments, sgRNAs were constructed by cloning the targeted sequences into the pX330 vector (Addgene #42230) via the BbsI sites. sgRNAs were designed using the GPP sgRNA designer from the BROAD institute. (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>). The expression of each sgRNA was driven by a U6 promotor. The same plasmid also expressed a human codon-optimized SpCas9 protein. The insertion donors were constructed based on the pUC57-Amp vector (Genewiz). Depending on the specific downstream experiments, the donors may be flanked by the same gene-targeting cutting site and/or by DRS-1 or DRS-2 sites (see Gao et al., 2019 for the latter two) (see Figure 1–figure supplement 3 for specific donor designs), which allows the donor to be excised by SpCas9 guided by the gene targeting sgRNA or by an additional DRS-1 or DRS-2 sgRNA. For some templates, a U6 promoter-driven DRS-2 sgRNA was integrated into the same plasmid of the donor to ensure co-expression with the donor and efficient excision during transfection. mRuby3 (Bajar et al., 2016) was expressed from a pKanCMV or a pCAGGS vector.

Cell culture and transfection

Human U2OS cells were ordered from ATCC (#HTB-96), and grown in McCoy's 5A medium (ATCC #30-2007) supplemented with 10% fetal bovine serum. Mouse Neuro 2A cells were originally from ATCC (#CCL-131) and were obtained as a gift from Dr. Kevin Wright's laboratory. Both cells were grown in a tissue culture incubator at 37°C and 5% CO₂. Cell culture transfections were carried out using Lipofectamine 2000 (ThermoFisher) following the standard protocol, except that only half of the recommended Lipofectamine amount (~5 µl/35 mm dish) was used with a lipofectamine to DNA ratio of 2 – 3. For confocal and SIM imaging, the cells were re-plated onto a cover-glass bottom dish (FluoroDish, WPI, #FD35-100) at 2 – 4 days post-transfection at an appropriate density and imaged after another 2 – 3 days. For Figure 4C, Neuro 2A cells were serum-starved (in 0.1% FBS) and treated with all-trans retinoic acid (ATRA, 12.5 µM) at 3 days post-transfection to induce cell differentiation, and then imaged after an additional 2 – 3 days.

FACS sorting and cell cloning

Transfected U2OS cells were subjected to fluorescence-activated cell sorting (FACS) at the OHSU flow cytometry core using an Influx sorter (BD). Red label-only (i.e., transfected but not successfully labeled) and green-labeled cells were collected separately. For cell cloning, 1 – 8 green-labeled cells were sorted into individual wells of a 96-well plate containing U2OS culture medium. Cell growth was visually inspected daily. Wells with only 1 – 3 dividing cells were monitored and later expanded into larger wells. A single-cell clone was used for Figure 6.

PCR and NGS

FACS-sorted cells (30k – 100k per reaction) were used. For genomic DNA preparation, cells were lysed in 0.05M NaOH and 0.1 mM EDTA at 95 °C for 15 minutes and then neutralized using 0.1 M (final concentration) Tris-HCl. RNA was isolated following an RNeasy Micro (Qiagen) protocol utilizing a QIAcube isolation robot at the OHSU Gene Profiling Shared Resource Core. mRNA was then reverse-transcribed using the GoScript reverse transcriptase enzyme (Promega) and random primers. Nested PCR was then carried out using OneTag DNA polymerase with GC buffer and Q5 polymerase (NEB) for genomic DNA and cDNA, respectively. GC buffer was necessary for genomic DNA amplification because the relevant intron contains highly GC-rich sequences. See Figure 2–figure supplement 1 and Supplementary Tables 1 and 2 for primer location, sequences, and combinations. A 4-bp barcode was added to the end of second-round primers to distinguish between individual samples. The first round of PCR for different products of each sample was carried out in a single reaction by including all relevant primers, and was amplified for 23 cycles with an annealing temperature of 60 °C. Approximately 2 µl of 10-fold diluted PCR products (to minimize the carryover of first-round primers) was added to 50 µl reactions for second-round PCRs with nested, specific primer pairs for 23 cycles. The exact added amount of first round PCR product was adjusted so that different samples gave comparable band intensities for the control PCR for *TUBA1B*. The PCR amplicons from different samples were mixed and submitted for NGS (Genewiz). The sequencing results were unmixed by the barcodes and by the presence of specific primers. INDELs were identified using custom code written in MATLAB.

Hippocampal slice culture and transfection

Hippocampi were dissected from P6 – P7 Sprague Dawley rat pups (both sexes) and sectioned to 400 μ m slices in dissection medium containing (in mM) 1 CaCl₂, 5 MgCl₂, 10 D-glucose, 4 KCl, 26 NaHCO₃, and 248 sucrose, with addition of 0.00025% phenol red. The slices were then seeded to a cell culture insert (Millipore, # PICMORG50) and cultured at 35°C with 5% CO₂ in medium containing 7.4 g/L MEM (ThermoFisher #11700-077) with the following supplements (in mM unless labeled otherwise): 16.2 NaCl, 2.5 L-Glutamax, 0.58 CaCl₂, 2 MgSO₄, 12.9 D-glucose, 5.2 NaHCO₃, 30 HEPES, 0.075% ascorbic acid, 1 μ g/ml insulin, and 20% heat-inactivated horse serum. Organotypic hippocampal slice cultures were transfected using the biolistic method. Plasmids were coated onto 1.6 μ m gold beads, and the slices were transfected at 4 – 7 days *in vitro* using a Helios gene gun (Bio-Rad). Slices were examined at 1 – 2 weeks post-transfection.

Animals, *in vivo* transfections, and sample preparations

Animal handling and experimental protocols were performed in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and were approved by the Institutional Animal Care and Use Committee (IACUC) of the Oregon Health & Science University (#IS00002792). *SpCas9* homozygous mice were crossed with C57BL/6 mice (Charles River). The resulting heterozygous offspring of both sexes were transfected using either IUE (at E16) or by stereotaxic viral injection (p14 – p60). In both cases, the somatosensory cortex was targeted. For *in vivo* imaging in living mice, a glass window was installed on the skull of the mouse at 1 – 4 months of age, as described previously (Ma et al., 2018). *In vivo* imaging was performed immediately after the window was installed.

IUE was performed at E16 as described previously (Ma et al., 2018) by injecting plasmid DNA (1 µl/embryo, total DNA concentration ~3-4 µg/µl, mixed with a 0.2% final concentration of fast green for visualization) into the lateral ventricle of mouse embryos, which were then electroporated with five 100-ms pulses (38V) using an electroporator (BEX #CUIY21). Stereotaxic viral injections were performed as previously described (Hunnicut et al., 2014) at ~p21 using a custom-modified David Kopf system. Typically, 20-50 nl of AAV was injected. For both IUE and viral injection, the somatosensory cortex was targeted.

Acute Brain Slices

Mice (p18 – 50) with neurons transfected using IUE were cardiac perfused with ice-cold, gassed (with 95% O₂ / 5% CO₂) artificial cerebral spinal fluid (ACSF) containing (in mM) 127 NaCl, 25 NaHCO₃, 25 D-glucose, 2.5 KCl, 1.25 NaH₂PO₄, 2 CaCl₂, and 1 MgCl₂. The brain was then dissected and coronal slices were prepared using a vibratome (Leica VT1200S) in a choline cutting solution (gassed with 95% O₂ / 5% CO₂) containing (in mM) 110 choline chloride, 25 NaHCO₃, 25 D-glucose, 2.5 KCl, 7 MgCl₂, 0.5 CaCl₂, 1.25 NaH₂PO₄, 11.5 sodium ascorbate, and 3 sodium pyruvate. The slices were the incubated in gassed ACSF at 35°C for 30 minutes and subsequently kept at room temperature for up to 6 hours.

Imaging and image analysis

All imaging experiments were carried out on live samples. Cultured cells were imaged using a custom-built two-photon microscope (up-right) (Ma et al., 2018), a Zeiss LSM 880 Airyscan laser-scanning confocal microscope, or an Elyra 7 lattice-based structured illumination microscope. Cultured and acute brain slices were imaged in a chamber perfused with gassed

ACSF. *In vivo* imaging was performed using a custom-built two-photon microscope based on the Janelia MIMMS design through a cranial window over transfected neurons in the mouse cortex. For two-photon microscopy, samples were excited using a Maitai HP Ti:Sapphire laser (Newport) at 960 or 990 nm, and green and red fluorescence were unmixed using a dichroic (Chroma 565DCXR) and band-pass filters (Chroma ET500/40 for green and Semrock FF01-630/92 for red). For Airyscan confocal microscopy and SIM, setups were configured according to the manufacturer's suggestions for Nyquist spatial sampling, and raw data was auto-processed. Throughout the manuscript, red fluorescence is presented as magenta.

Image analyses were performed using custom software written in MATLAB. Labeling efficiencies were calculated as successful CRISPIEd cells (green) divided by total transfected cells (red) in random field of views (FOVs) (2p imaging; 330 X 330 or 165 X 165 μ m FOV) selected based on the red channel only. For the relative labeling efficiency across constructs in Figure 3, the numbers of total CRISPIEd (green) cells were visually counted from nine FOVs of predetermined arrangement (a 3 X 3 grid separated by 2.5 mm) under a 10X objective. All conditions to be compared were transfected and imaged side-by-side, and data were normalized to the averaged cell counts per FOV of the best labeled transfections. For IUE experiments, only the centers of the transfected regions were imaged and quantified. For mTurquoise2 and mNeonGreen labeling, the images were taken under the same conditions as mEGFP but were pseudo-colored in the color of turquoise and neon green, respectively.

Spine enrichment index (SEI) measurements were calculated as described previously (Zhong et al., 2009). Briefly, averaged green and red intensities were measured from manually

drawn regions of interest (ROIs) on the spines and their adjacent parental dendritic shaft in 2p images. SEI was then calculated as $\log_2[(F_{\text{green}}/F_{\text{red}})_{\text{spine}}/(F_{\text{green}}/F_{\text{red}})_{\text{shaft}}]$, where F is the average fluorescence intensities of individual ROIs. Positive values indicate relative spine enrichment, whereas negative values indicate relative spine exclusion.

Sample size, replication, and sample allocation

Sample sizes, as indicated in figure legends, were determined based on the observed variability across measurements. All experiments have been replicated in multiple independent trials/animals (typically >3) with the exception that only one round of NGS was carried out. Cells and animals were randomly allocated to each experimental groups. All groups to be compared were carried out side-by-side. The researchers were not blinded to group allocation during data collection and analysis. However, the field of views were selected either predetermined or in the absence of knowledge of the color channel under investigation.

Quantification, presentation, and statistics

Quantification and statistical tests were performed using MATLAB scripts. p values were obtained from one-way ANOVA tests, unless noted otherwise. In all figures, *: $p \leq 0.05$ and is statistically significant after Bonferroni correction for multiple tests, **: $p \leq 0.01$, and ***: $p \leq 0.001$.

AUTHOR CONTRIBUTIONS

H.Z. conceived the project and designed the experiments with inputs from T.M. H.Z. carried out the majority of experiments with help on specific aspects from C.I.M., M.A.M., L.M., M.Q., and S.K.P. The actin dynamics experiment was proposed and carried out by S.K.P. H.Z. analyzed the data. H.Z. and T.M. wrote the manuscript with edits and comments from all authors. H.Z. and T.M. secured the funding.

ACKNOWLEDGEMENTS

We thank Dr. Zefeng Wang for discussions regarding exon-intron junctions; Drs. Paul Meraner and Marc Freeman for the *SpCas9* mouse line; Dr. Kevin Wright for the Neuro 2A cell line; Dr. Bart Jongbloets for help with imaging, Dr. James Frank for help with experimental attempts that are not included, Mr. Matthew Schleisman at the OHSU flow cytometry core and Ms. Brittany Daughtry at the OHSU Gene Profiling Shared Resource for technical assistance on FACS sorting and mRNA purification experiments, respectively. We thank Dr. Gail Mandel for critical comments on the manuscript. The work was supported by a NIH/BRAIN Initiative grant (RF1MH120119) to H.Z. and T.M., and an NIH/NINDS R01 (R01NS081071) to T.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. doi:10.1038/s41586-019-1711-4
- Bajar BT, Wang ES, Lam AJ, Kim BB, Jacobs CL, Howe ES, Davidson MW, Lin MZ, Chu J. 2016. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci Rep*. doi:10.1038/srep20889
- Chen B, Zou W, Xu H, Liang Y, Huang B. 2018. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nat Commun*. doi:10.1038/s41467-018-07498-y
- Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA. 2017. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*. doi:10.1038/nature24268
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* (80-). doi:10.1126/science.1231143
- Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. doi:10.1038/nbt.3437

525 Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier
526 RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene
527 inactivation. *Nat Biotechnol*. doi:10.1038/nbt.3026

528 Doudna JA. 2020. The promise and challenge of therapeutic genome editing. *Nature*.
529 doi:10.1038/s41586-020-1978-5

530 Fischer M, Kaech S, Knutti D, Matus A. 1998. Rapid actin-based plasticity in dendritic spines.
531 *Neuron*. doi:10.1016/S0896-6273(00)80467-5

532 Fortin DA, Tillo SE, Yang G, Rah JC, Melander JB, Bai S, Soler-Cedeño O, Qin M, Zemelman B V.,
533 Guo C, Mao T, Zhong H. 2014. Live imaging of endogenous PSD-95 using ENABLED: A
534 conditional strategy to fluorescently label endogenous proteins. *J Neurosci*.
535 doi:10.1523/JNEUROSCI.3888-14.2014

536 Gao Y, Hisey E, Bradshaw TWA, Erata E, Brown WE, Courtland JL, Uezu A, Xiang Y, Diao Y,
537 Soderling SH. 2019. Plug-and-Play Protein Modification Using Homology-Independent
538 Universal Genome Engineering. *Neuron*. doi:10.1016/j.neuron.2019.05.047

539 Goldie BJ, Cairns MJ. 2012. Post-transcriptional trafficking and regulation of neuronal gene
540 expression. *Mol Neurobiol*. doi:10.1007/s12035-011-8222-0

541 He M, Huang ZJ. 2018. Genetic approaches to access cell types in mammalian nervous systems.
542 *Curr Opin Neurobiol*. doi:10.1016/j.conb.2018.02.003

543 Heidenreich M, Zhang F. 2016. Applications of CRISPR-Cas systems in neuroscience. *Nat Rev*

544 *Neurosci.* doi:10.1038/nrn.2015.2

545 Herzog E, Nadrigny F, Silm K, Biesemann C, Helling I, Bersot T, Steffens H, Schwartzmann R,
546 Valentin Nägerl U, El Mestikawy S, Rhee JS, Kirchhoff F, Brose N. 2011. In vivo imaging of
547 intersynaptic vesicle exchange using VGLUT1 Venus knock-in mice. *J Neurosci.*
548 doi:10.1523/JNEUROSCI.2073-11.2011

549 Heyer WD, Ehmsen KT, Liu J. 2010. Regulation of homologous recombination in eukaryotes.
550 *Annu Rev Genet.* doi:10.1146/annurev-genet-051710-150955

551 Hunnicutt BJ, Long BR, Kusefoglu D, Gertz KJ, Zhong H, Mao T. 2014. A comprehensive
552 thalamocortical projection map at the mesoscopic level. *Nat Neurosci.*
553 doi:10.1038/nn.3780

554 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-
555 RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* (80-).
556 doi:10.1126/science.1225829

557 Kaessmann H, Zöllner S, Nekrutenko A, Li WH. 2002. Signatures of domain shuffling in the
558 human genome. *Genome Res.* doi:10.1101/gr.520702

559 Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. 2016. High-
560 fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.*
561 doi:10.1038/nature16526

562 Lam AJ, St-Pierre F, Gong Y, Marshall JD, Cranfill PJ, Baird MA, McKeown MR, Wiedenmann J,

Davidson MW, Schnitzer MJ, Tsien RY, Lin MZ. 2012. Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat Methods*. doi:10.1038/nmeth.2171

Lee PT, Zirin J, Kanca O, Lin WW, Schulze KL, Li-Kroeger D, Tao R, Devereaux C, Hu Y, Chung V, Fang Y, He Y, Pan H, Ge M, Zuo Z, Housden BE, Mohr SE, Yamamoto S, Levis RW, Spradling AC, Perrimon N, Bellen HJ. 2018. A gene-specific T2A-GAL4 library for drosophila. *Elife*. doi:10.7554/eLife.35574

Li Jun, Meng X, Zong Y, Chen K, Zhang H, Liu J, Li Jiayang, Gao C. 2016. Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nat Plants*. doi:10.1038/nplants.2016.139

Ma L, Jongbloets BC, Xiong WH, Melander JB, Qin M, Lameyer TJ, Harrison MF, Zemelman B V., Mao T, Zhong H. 2018. A Highly Sensitive A-Kinase Activity Reporter for Imaging Neuromodulatory Events in Awake Mice. *Neuron*. doi:10.1016/j.neuron.2018.07.020

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science (80-)*. doi:10.1126/science.1232033

Mao Z, Bozzella M, Seluanov A, Gorbunova V. 2008. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*. doi:10.4161/cc.7.18.6679

Mikuni T, Nishiyama J, Sun Y, Kamasawa N, Yasuda R. 2016. High-Throughput, High-Resolution Mapping of Protein Localization in Mammalian Brain by in Vivo Genome Editing. *Cell*. doi:10.1016/j.cell.2016.04.044

583 Miura H, Gurumurthy CB, Sato T, Sato M, Ohtsuka M. 2015. CRISPR/Cas9-based generation of
584 knockdown mice by intronic insertion of artificial microRNA using longer single-stranded
585 DNA. *Sci Rep*. doi:10.1038/srep12799

586 Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, Daimon T, Sezutsu H, Yamamoto
587 T, Sakuma T, Suzuki KIT. 2014. Microhomology-mediated end-joining-dependent
588 integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat*
589 *Commun*. doi:10.1038/ncomms6560

590 Nakamura M, Srinivasan P, Chavez M, Carter MA, Dominguez AA, La Russa M, Lau MB, Abbott
591 TR, Xu X, Zhao D, Gao Y, Kipniss NH, Smolke CD, Bondy-Denomy J, Qi LS. 2019. Anti-CRISPR-
592 mediated control of gene editing and synthetic circuits in eukaryotic cells. *Nat Commun*.
593 doi:10.1038/s41467-018-08158-x

594 Nishiyama J, Mikuni T, Yasuda R. 2017. Virus-Mediated Genome Editing via Homology-Directed
595 Repair in Mitotic and Postmitotic Cells in Mammalian Brain. *Neuron*.
596 doi:10.1016/j.neuron.2017.10.004

597 Patthy L. 1999. Genome evolution and the evolution of exon-shuffling - A review. *Gene*.
598 doi:10.1016/S0378-1119(99)00228-0

599 Planchon TA, Gao L, Milkie DE, Davidson MW, Galbraith JA, Galbraith CG, Betzig E. 2011. Rapid
600 three-dimensional isotropic imaging of living cells using Bessel beam plane illumination.
601 *Nat Methods*. doi:10.1038/nmeth.1586

602 Rizzo MA, Davidson MW, Piston DW. 2009. Fluorescent protein tracking and detection:

603 Fluorescent protein structure and color variants. *Cold Spring Harb Protoc.*
604 doi:10.1101/pdb.top63

605 Roberts B, Haupt A, Tucker A, Grancharova T, Arakaki J, Fuqua MA, Nelson A, Hookway C,
606 Ludmann SA, Mueller IA, Yang R, Horwitz R, Rafelski SM, Gunawardane RN. 2017.
607 Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell
608 organization. *Mol Biol Cell.* doi:10.1091/mbc.E17-03-0209

609 Saleh-Gohari N, Helleday T. 2004. Conservative homologous recombination preferentially
610 repairs DNA double-strand breaks in the S phase of the cell cycle un human cells. *Nucleic*
611 *Acids Res.* doi:10.1093/nar/gkh703

612 Schmid-Burgk JL, Höning K, Ebert TS, Hornung V. 2016. CRISPaint allows modular base-specific
613 gene tagging using a ligase-4-dependent mechanism. *Nat Commun.*
614 doi:10.1038/ncomms12338

615 Shroff H, Galbraith CG, Galbraith JA, Betzig E. 2008. Live-cell photoactivated localization
616 microscopy of nanoscale adhesion dynamics. *Nat Methods.* doi:10.1038/nmeth.1202

617 Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9
618 nucleases with improved specificity. *Science (80-).* doi:10.1126/science.aad5227

619 Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, Hatanaka F, Yamamoto M,
620 Araoka T, Li Z, Kurita M, Hishida T, Li M, Aizawa E, Guo S, Chen S, Goebel A, Soligalla RD, Qu
621 J, Jiang T, Fu X, Jafari M, Esteban CR, Berggren WT, Lajara J, Nuñez-Delicado E, Guillen P,
622 Campistol JM, Matsuzaki F, Liu GH, Magistretti P, Zhang Kun, Callaway EM, Zhang Kang,

623 Belmonte JCI. 2016. In vivo genome editing via CRISPR/Cas9 mediated homology-
624 independent targeted integration. *Nature*. doi:10.1038/nature20565

625 Uemura T, Mori T, Kurihara T, Kawase S, Koike R, Satoga M, Cao X, Li X, Yanagawa T, Sakurai T,
626 Shindo T, Tabuchi K. 2016. Fluorescent protein tagging of endogenous protein in brain
627 neurons using CRISPR/Cas9-mediated knock-in and in utero electroporation techniques.
628 *Sci Rep*. doi:10.1038/srep35861

629 Wang Z, Burge CB. 2008. Splicing regulation: From a parts list of regulatory elements to an
630 integrated splicing code. *RNA*. doi:10.1261/rna.876308

631 Willems J, de Jong APH, Scheefhals N, Mertens E, Catsburg LAE, Poorthuis RB, de Winter F,
632 Verhaagen J, Meye FJ, MacGillavry HD. 2020. Orange: A CRISPR/Cas9-based genome
633 editing toolbox for epitope tagging of endogenous proteins in neurons. *PLoS Biol*.
634 doi:10.1371/journal.pbio.3000665

635 Zhong H, Sia GM, Sato TR, Gray NW, Mao T, Khuchua Z, Haganir RL, Svoboda K. 2009.
636 Subcellular Dynamics of Type II PKA in Neurons. *Neuron*.
637 doi:10.1016/j.neuron.2009.03.013

638

(B) Schematics of the sgRNA/SpCas9 plasmid (upper) and the donor plasmid (lower) that are used for panels D–G. Purple triangles: the sgRNA targeted location and orientation, spl.: splicing; acptr: acceptor. See Figure 1–figure supplement 3 for design details.

(C) Schematic of the targeted intron of *hACTB* with the purple triangle showing the sgRNA targeting site and orientation. Orange and gray boxes: exonic sequences that do and do not encode protein sequences, respectively.

(D) Representative two-photon (2p) images of live U2OS cells with the indicated transfection. Note that the mRuby3 (magenta) expression levels were highly variable across cells due to the variability in plasmid transfection; however, the green label intensities were comparable across cells, as expected for the expression from an endogenous locus.

(E) Labeling efficiency (green cell counts over red cell counts) for *hACTB* in U2OS cells. $n = 8$ field of views (FOVs), two independent transfections.

(F) Representative confocal (Airyscan) and super-resolution SIM images of live U2OS cells expressing CRISPIEd β -actin.

(G) Representative time-lapse confocal images of live U2OS cells showing the dynamics of actin ruffles (arrows). The dashed yellow line outlines the cell morphology at time zero.

See also:

Figure 1–figure supplement 1. DNA insertion targeting coding sequences results in many possible modes of unwanted mutations.

666 Figure 1—figure supplement 2. Introns are much less conserved than exons.

667 Figure 1—figure supplement 3. Schematic of targeting sites and donor designs.

668 Figure 1—Source Data 1. Numeric data for Figure 1E.

36

(B) Gel images of PCR and RT-PCR analyses for the genomic DNA and mRNA, respectively, of successfully labeled cells to detect non-labeled events, and both 5' and 3' junctions of forward and inverted label insertion events.

(C) Representative NGS sequencing results of the genomic DNA and mRNA prepared from transfected U2OS cells without successful actin labeling corresponding to bands ii and ii', respectively. INDELs are marked in red.

(D) Pie graphs of the relative proportions of INDEL events based on NGS results of the PCR/RT-PCR products from panels A and B, as marked by the Roman numerals. Statistical tests were performed using a χ^2 test. From top to bottom, n = 54216, 1181, 1589, 2767, and 6036 for genomic DNA, and 5771, 2936, 4166, 3613, and 5949 for mRNA.

See also:

Figure 2—figure supplement 1. Schematic of PCR and RT-PCR products.

Figure 2—Source Data 1. Numeric data for Figure 2D.

Figure 3. CRISPIE is optimizable and achieves higher efficiency than exon labeling.

(A) Schematic positions of the tested editing sites in the intron 1 and exon 2 of *hACTB* that are targeted in panels (B) and (C). The asterisk marks the site used for *hACTB* in subsequent experiments.

(B) Comparison of the editing efficiency of five different intronic locations. For comparison under identical conditions, a generic donor excised using an independent sgRNA (DRS-2) was used (Figure 1–figure supplement 3B7). n = 18 FOVs from two independent transfections.

(C) Representative images and quantification of successful editing rates at the intronic location i4 versus the only two possible exonic locations for N-terminally labeled *hACTB*, each using their specifically designed donors. n = 18 FOVs from two independent transfections.

See also:

Figure 3–Source Data 1. Numeric data for Figures 3B and 3C.

Figure 4. CRISPIE is applicable to labeling with different colors, different proteins, and different animal species.

(A) Representative images of hACTB labeled with each of the three indicated FPs of different colors in U2OS cells.

(B) Representative images of four different proteins, as indicated, labeled with mEGFP in U2OS cells.

(C) Representative images of hACTB and mACTB labeling in human U2OS cells and mouse Neuro2A cells, respectively, under the indicated conditions.

Figure 5. CRISPIE is applicable to different proteins in postmitotic neurons using various transfection methods.

(A) Representative images of the indicated proteins labeled in rat cultured hippocampal slices transfected using the biolistic method.

(B) Representative images of the indicated proteins labeled in cultured hippocampal slices from *SpCas9* +/- mice transfected using the biolistic method. Arrowheads indicate spines where each of the three proteins are differentially enriched or excluded.

(C) SEI of the three proteins in panel (B). From left to right, n (spines/neurons) = 55/5, 26/3, and 22/2.

(D, E) Representative images (D) and labeling efficiency quantifications (E) of β -actin or α -tubulin 1B-labeled neurons that are labeled *in vivo* via IUE in the mouse somatosensory cortex and imaged in acute brain slices. n = 10 FOVs from three mice for the quantification of both proteins.

(F) Representative side view (x-z, left), top view (x-y, middle), and zoomed-in (x-y, right) images of labeled mACTB (green) in cortical neurons that were transfected via IUE and were imaged in living mice via a cranial window.

(G) Representative images of β -actin-CRISPIEd neurons in acute brain slices labeled via AAV injections into the cortex of *SpCas9 +/-* mice.

See also:

Figure 5—figure supplement 1. The CRISPIE method is suitable for labeling neurons in cultured hippocampal slices transfected using the biolistic (gene gun) method.

Figure 5—figure supplement 2. CRISPIEd mACTB transfected using *in utero* electroporation is suitable for *in vivo* imaging.

Figure 5—Source Data 1. Numeric data for Figures 5C and 5E.

Figure 6. Labeling with CRISPIE can be erased.

(A) Schematic of the experimental design for erasing CRISPIE labels. DRS-1 sgRNA targeting sequences were used to flank the exon/intron module inside the other two sgRNA sites for excising the donor from the initial insertion. Later introduction of DRS-1 sgRNA and SpCas9 can excise the inserted CRISPIE module. Note that after the insertion and erasure, both the original sgRNA targeting site and the DRS-1 site will be destroyed.

(B) Representative images of a β -actin-CRISPIEd U2OS cell clone (left), which was transfected with the transfection marker mRuby3 (magenta in upper panels) and SpCas9 together with

748 DRS-1 (middle), or with a control sgRNA that targeted mACTB, but not hACTB (right). Asterisks
749 in the lower panel mark the transfected (i.e., mRuby-positive) cells.

750 (C) Quantification of the erasure efficiency among the transfected cells, as identified by
751 mRuby3 expression. n (FOVs/transfections) = 11/3 for DRS-1, and 9/3 for control sgRNA.

752 See also:

753 Figure 6—Source Data 1. Numeric data for Figure 6C.

754

Figure 1—figure supplement 1. DNA insertion targeting coding sequences results in many possible modes of unwanted mutations.

Schematic of the design and consequences of DNA insertions targeting coding sequences.

Orange boxes: endogenous exons; green boxes: coding sequences to be inserted, with arrows indicating the orientation of the reading frame; red stars: INDELS. The blue check mark indicates desired precise insertion, while red X's indicate unwanted mutant mRNAs.

763

764 Figure 1–figure supplement 2. Introns are much less conserved than exons.

765 Representative alignment of human (h) and mouse (m) *ACTB* at the junctions of the two ends of

766 its first intron. Identical base pairs are highlighted in gray.

767

Figure 1—figure supplement 3. Schematic of targeting sites and donor designs.

(A) Schematic of the genomic structure around the targeting sites. Short vertical purple lines indicate the site positions and whether the targeted sequences are at the upper (purple lines above the DNA schematic) or lower (purple lines below the DNA schematic) DNA strand. All scale bars are 250 bp. (B) Schematic donor plasmid designs used in this manuscript. Triangles illustrate the indicated sgRNA targeting sites with orientation. Note that CRISPIE allows for the inclusion of multiple cutting sites, as indicated, which provides flexibility for excising the donor DNA or the possibility to erase the inserted module at a later time. The phases of the exon

777 modules (i.e., the reading frames) are labeled. The junctional intron (100 bp) and exon (~10 bp)
 778 sequences are taken from intron 17 of mouse *CaMKII α* and its surrounding exons, with the
 779 exception that the 5' intron (250 bp)/exon junction of b5 is taken from intron 21 and exon 22 of
 780 *hVCL*, b6 does not involve any intron/exon junction (targeting coding sequences, or CDS), and
 781 the 3' intron (250 bp)/exon junction of b8 is taken from exon 18 and intron 19 of mouse *PSD-*
 782 *95*. The donors of b1 – b5 must be released from the parental plasmid by an additional
 783 sgRNA/SpCas9 (not depicted here) that cuts one of the indicated flanking sgRNA targeting sites.
 784 b6 – b10 can be excised using the DRS-2 sgRNA from the same plasmid in the presence of
 785 SpCas9. (C) Schematic AAV for use with SpCas9-expressing cells/animals.

786

Figure 2–figure supplement 1. Schematic of PCR and RT-PCR products.

The thick black lines illustrate the PCR and RT-PCR products corresponding to those in Figure 2. Orange and gray boxes: exonic sequences that do and do not encode protein sequences, respectively; thin black lines: endogenous intron sequences; thick blue lines and green boxes: intron and exon sequences, respectively, of the designer donor module with white arrows indicating the correct orientation of the reading frame. Thick black dashed line and the light green box enclosed by dashed lines illustrate the hypothetical PCR products and the inverted EGFP exon, respectively, for inversely inserted mRNA. Note that this condition is only hypothetical and should not exist because there is no appropriate splicing acceptor and donor sites when the donor module is inserted in the inverted orientation.

799

800 Figure 5-figure supplement 1. The CRISPIE method is suitable for labeling neurons in cultured
801 hippocampal slices transfected using the biolistic (gene gun) method.

802 (A) Representative macroscopic image showing multiple CRISPIEd neurons (mEGFP-labeled
803 rACTB) in one cultured rat hippocampal slice. (B) Representative two-photon images of
804 CRISPIEd rACTB of various cell types in cultured rat hippocampal slices. The CA1 and CA3

805 neurons are composites of multiple FOVs. The CA3 neuron is the same as the one presented in
806 the left panel of Figure 5A.

807

808 Figure 5—figure supplement 2. CRISPIEd mACTB transfected using *in utero* electroporation is
 809 suitable for *in vivo* imaging.

810 Representative x-z view of a two-photon image taken through a cranial window in a living
 811 mouse transfected using *in utero* electroporation at E16.

812

Primer #	Sequence
P#1	CCGCGAAGCCGGCCTTGAC
P#2	GAGCACAGAGCCTCGCCTTTGC
P#3	GAGCCTCGCCTTTGCCGATCCG
P#4	GCCAGATTTTCTCCATGTCGTCCC
P#5	CACCATCACGCCCTGGTGCC
P#6	GACCCGGCGCTGTTTGAACC
P#7	GCTGTTTGAACCGGGCGGAGG
P#8	CATCGTCGCCCCGGAAGCC
P#9	GGTGCATTAAGGGTATTGCCTTAG
P#10	GTGAGGGTGGAGGACCTGTAC
P#11	GTTTACGTCGCCGTCCAGCTC
P#12	CACCACCCCGGTGAACAGCTC
P#13	GCGGACTTGAAGAAGTCGTGC
P#14	GCTTCATGTGGTCGGGGTAGC
P#15	CCTGCTGGAGTTCGTGACCG
P#16	CGAGAAGCGCGATCACATGG
P#17	CTGCTCCTGTCGCCTTCG
P#18	CGCCTCCTAATCCCTAGCCAC
P#19	GACTGCGGGGACTCGAGG
P#20	CCCAGGACACAGTTACGCG
P#21	GTGTAGTGCCCTCGGGCATAG
P#22	GCAGCATCTTCCTTGCCTGTG

Supplementary Table 1. PCR primer list.

		First round primers		Second round (nested) primers		Expected size (bp)
ACTB Genomic DNA	No donor insertion	P#6	P#8	P#7	P#1	325
	Donor insertion, forward, 5'	P#6	P#11	P#7	P#12	314
	Donor insertion, forward, 3'	P#9	P#8	P#10	P#1	324
	Donor insertion, inverse, 5'	P#6	P#9	P#7	P#10	179
	Donor insertion, inverse, 3'	P#11	P#8	P#12	P#1	459
ACTB mRNA	No donor insertion	P#2	P#4	P#3	P#5	187
	Donor insertion, forward, 5'	P#2	P#13	P#3	P#14	308
	Donor insertion, forward, 3'	P#16	P#4	P#15	P#5	247
	Donor insertion, inverse, 5'	P#2	P#16	P#3	P#15	n/a
	Donor insertion, inverse, 3'	P#13	P#4	P#14	P#5	n/a
TUBA1B	Genomic DNA, ctrl	P#17	P#19	P#18	P#20	324
	mRNA, ctrl	P#17	P#21	P#18	P#22	321

817 Supplementary Table 2. PCR primer combinations.

818